

AD-A158 645

A FIVE-PARAMETER FAMILY OF PROBABILITY DISTRIBUTIONS
REVISION 1(U) PURDUE UNIV LAFAYETTE IN SCHOOL OF
INDUSTRIAL ENGINEERING B W SCHMELSER ET AL. AUG 85
RM-84-7-REV-1 N00014-79-C-0832

1/1

UNCLASSIFIED

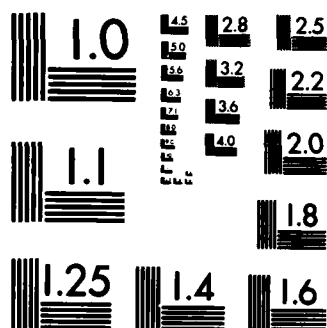
F/G 12/1

NL

END

FILED

STOP



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

2

AD-A158 645

A FIVE-PARAMETER FAMILY OF PROBABILITY DISTRIBUTIONS

Bruce W. Schmeiser
School of Industrial Engineering
Purdue University
West Lafayette, Indiana 47907

Ram Lal
Department of Management Science
California State University
Fullerton, CA 92634

DTIC
ELECTE
SEP 4 1985
S B D

Technical Report 84-7
School of Industrial Engineering
Purdue University

April 1984
Revised August 1985

Research supported by Office of Naval Research contracts N00014-77-C-0425 and
N00014-79-C-0832.

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited

ABSTRACT

A tractable, five-parameter family of continuous, unimodal probability distributions is developed. Special cases include the Bernoulli trial, uniform, power series, exponential, triangular and Laplace (double exponential) distributions. The family has closed-form density function, cumulative distribution function, inverse cumulative distribution function, hazard function, and residual moments. Statistical properties, parameter determination and random variate generation are discussed. Two numerical examples are given.

Additional Keywords:

method;

KEY WORDS: Input modelling; Monte Carlo *method;* Process generation, Simulation, Statistical modelling.

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail. and/or Special
A-1	



1. INTRODUCTION

We consider the five-parameter family of probability distributions having density function

$$f(x) = \begin{cases} [1 + (x-\alpha)(\delta/\epsilon)^{1/2}/(\beta\gamma)]^{(1-\delta)/\delta} / (\beta(\delta\epsilon)^{1/2}) & \text{if } a \leq x \leq \alpha \\ [1 - (x-\alpha)(\epsilon/\delta)^{1/2}/(\beta(1-\gamma))]^{(1-\epsilon)/\epsilon} / (\beta(\delta\epsilon)^{1/2}) & \text{if } \alpha \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $a = \alpha - \beta\gamma(\epsilon/\delta)^{1/2}$ is the lower bound and $b = \alpha + \beta(1-\gamma)(\delta/\epsilon)^{1/2}$ is the upper bound of the random variable X . The parameters have ranges $-\infty < \alpha < \infty$, $0 < \beta < \infty$, $0 \leq \gamma \leq 1$, $0 < \delta < \infty$, and $0 < \epsilon < \infty$. This family can model a wide variety of univariate, continuous, unimodal probabilistic phenomena by varying the five parameters. Properties of the family are developed in Section 2, parameter determination is discussed in Section 3, and two numerical examples are given in Section 4. The rest of this section is a brief review of related literature.

Schmeiser (1977) surveys the literature of versatile families of distributions, including the Pearson and Johnson systems of distributions; approximations to the inverse distribution function via expansion techniques, polynomial regression, and rectangular approximation; and various four parameter distributions. Johnson and Kotz (1970) review various classical methods in some detail. We focus here on the more recent literature, for which the motivation often has been to develop families useful as input models in Monte Carlo simulation.

Johnson et al. (1980) develop a four-parameter family of symmetric distributions that includes the uniform, normal, and Laplace distributions as special cases. Since random variate generation is a simple transformation of a gamma random variate and since the kurtosis is a simple function of the parameters, the family is ideal for Monte

Carlo studies of the effects of departures from normality. Fitting the distribution to data is not discussed.

Stacy (1962), Stacy and Mlhrum (1965), and Harter (1967) discuss the four-parameter gamma distribution, which is based on taking a gamma random variable to a power. Both the three-parameter Weibull and gamma distributions are special cases. This distribution is useful for modeling lifetimes, although parameter estimation is not straightforward.

Hora (1983) develops a modeling procedure based on the inverse distribution function. The analyst selects a tractable distribution function F that is a reasonable approximation. The model is $F^{-1}(g(p))$, where $g(p)$ is a polynomial model fitted by regression to the data. Random variate generation and data fitting are straightforward, but the analyst must provide the distribution function F . The families discussed below provide good candidates for F .

Burr (1942, 1973) developed a more flexible distribution with the same range and straightforward random variate generation. Heavy and light tailed distributions are not obtainable, nor are symmetric distributions, but the family is useful for modeling lifetimes, although again parameter estimation is difficult.

Ramberg and Schmeiser (1972, 1974) generalized Tukey's (1960) lambda distribution to obtain a family for all but light tailed distributions, including symmetric distributions. The exponential distribution is a limiting case and the normal distribution function is approximated within a tolerance of .001 in the distribution function. (See Joiner and Rosenblatt (1971) for a related approximation.) Random variate generation is straightforward. Dudewicz et al. (1974) and Ramberg et al. (1979) provide tables to fit the distribution by matching moments. Other estimation techniques, such as maximum likelihood estimation, are not straightforward due to the

distribution function and density function not being expressible in closed form, although Öztürk and Dale (1985) consider minimizing the squared errors between the empirical and fitted distribution functions.

Schmeiser and Deutsch (1977) describe a family of distributions that can be used to obtain a distribution having any first four moments. The exponential distribution is a limiting case. Closed form expressions are given to calculate parameter values to match the desired mode, a percentile of the mode and any other two quantiles. Moments may be matched graphically or via nonlinear programming. Random variate generation is closed form and requires only one exponentiation operation. Kottas and Lau (1979) use the distribution in inventory modeling as the distribution of lead-time demand and Lau and Zaki (1982) use the distribution to test sensitivity to lead-time-demand assumptions. The disadvantage of this distribution is that while any four independent properties may be specified, the shape of the distribution is only satisfactory as a rough approximation. The shortcomings involve truncated tails and a density function value at the mode that is often either zero or infinity. This family could be generalized, with little loss of tractability, to five parameters analogous to the family developed here.

The family of distributions developed here has continuous densities, obtains any mean and variance as well as a wide range of commonly used third and fourth moments, has straightforward random variate generation, has closed-form moments, and has tractable parameter determination.

2. PROPERTIES OF THE FAMILY

As follows directly from the density function, the distribution function is

$$F(x) = \begin{cases} 0 & \text{if } -\infty < x < a \\ \gamma[1 + (x-\alpha)(\delta/\epsilon)^{1/2} / (\beta\gamma)]^{1/\delta} & \text{if } a \leq x \leq \alpha \\ 1 - \{(1-\gamma)[1 - (x-\alpha)(\epsilon/\delta)^{1/2} / (\beta(1-\gamma))]\}^{1/\epsilon} & \text{if } \alpha < x \leq b \\ 1 & \text{if } b < x < \infty \end{cases}$$

where again $a = \alpha - \beta\gamma(\epsilon/\delta)^{1/2}$ and $b = \alpha + \beta(1-\gamma)(\delta/\epsilon)^{1/2}$ are the lower and upper bounds.

An appealing feature of this family of distributions is that the role of each of the five parameters is relatively intuitive and independent of the others. The location parameter, α , is the mode or antimode. The scale parameter, β , is the range when $\delta = \epsilon$ and is always proportional to the standard deviation, which is also a function of γ , δ , and ϵ . These three remaining parameters determine shape, but again each has a specific role. The probability to the left of the mode, a rough measure of skewness, is γ ; i.e., $P\{X \leq \alpha\} = \gamma$. The shape of the distribution is determined primarily by δ to the left of α and by ϵ to the right of α . The shape is symmetric if and only if $\delta = \epsilon$ and $\gamma = 1/2$. For $\delta > 1$ ($\epsilon > 1$), the curve to the left (right) of the mode will be J-shaped with skew to the right (left). For $\delta < 1$ ($\epsilon < 1$), the curve to the left (right) of the mode will be J-shaped with skew to the left (right).

These parameters yield a wide range of shapes, as can be seen by considering some special cases. First consider symmetric distributions, for which $\delta = \epsilon$ and $\gamma = 1/2$. As $\delta = \epsilon$ approaches infinity with $\alpha = \gamma = 1/2$ and $\beta = 1$, $P\{X = 0\} \rightarrow 1/2$ and $P\{X = 1\} \rightarrow 1/2$, a fair Bernoulli trial. The uniform distribution over the interval $[a, b]$ is obtained by setting $\alpha = a + (b-a)\gamma$, $\beta = b-a$, $\delta = \epsilon = 1$. (In this one case, γ

may assume any value in the unit interval and the distribution remains symmetric.) As $\delta = \epsilon$ approaches zero with $\gamma = 1/2$, the limiting distribution is the Laplace. Thus for symmetric distributions the family provides a kurtosis range from one (Bernoulli trial) to six (Laplace). Very heavy tailed distributions cannot be obtained.

A wide range of asymmetric shapes is also possible. As $\delta = \epsilon$ approaches infinity with $\alpha = \gamma$ and $\beta = 1$, $P\{X = 0\} \rightarrow \gamma$ and $P\{X = 1\} \rightarrow 1 - \gamma$, a biased Bernoulli trial. By letting $\alpha = 0$, $\beta = \lambda/\epsilon$, $\gamma = 0$, and $\delta = \epsilon$, the limiting case as $\epsilon \rightarrow 0$ is exponential with mean λ , as can be seen from by considering nonnegative values of x in

$$\lim_{\epsilon \rightarrow 0} F(x) = \lim_{\epsilon \rightarrow 0} 1 - [1 - x/(\lambda/\epsilon)]^{1/\epsilon} = 1 - e^{-x/\lambda}$$

which is the distribution function of the exponential distribution having mean λ . The power series distribution having density function

$$f(x) = \begin{cases} \lambda x^{\lambda-1} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

is obtained when $\alpha = \beta = \gamma = 1$ and $\delta = \epsilon = 1/\lambda$.

Some graphs illustrate the versatility of the family. Figure 1 shows graphs of the distribution for $\gamma = 0.5$ for various values of $\delta = \epsilon$. Changing α and/or β would change the location and/or scale, but not the shape, of the distribution. Since $\gamma = 0.5$, the distribution is symmetric for all $\delta = \epsilon$. Figure 2 shows graphs corresponding to $\gamma = 0.75$. Here 75% of the area under each curve lies to the left of the mode (anti-mode) α for all $\delta = \epsilon$. When $\delta \neq \epsilon$, the left and right sides of the distribution differ, resulting in arbitrary combinations of tail shapes.

Solving for x as a function of p in the distribution function yields the inverse distribution function

$$x = F^{-1}(p) = \begin{cases} \alpha - \beta \gamma (\epsilon/\delta)^{1/2} [1 - (p/\gamma)^\delta] & \text{if } 0 \leq p \leq \gamma \\ \alpha + \beta(1-\gamma)(\delta/\epsilon)^{1/2} [1 - ((1-p)/(1-\gamma))^\epsilon] & \text{if } \gamma < p \leq 1 \end{cases} \quad (2)$$

which gives the p^{th} quantile of X as a function of p . In addition to calculating quantiles, the inverse transformation is important as a method of random variate generation. If $P \sim U(0,1)$, then $F^{-1}(P) \sim X$. Random variate generation based on (2) is relatively fast in that only one exponentiation is required. The closed-form inverse distribution function allows the direct generation of order statistics via the methods of Lurie and Hartley (1972), Schmelser (1978a), Ramberg and Tadikamalla (1978), and Schucany (1972), as surveyed in Schmelser (1978b). The inverse transformation is also important in simulation experiments as a method to induce positive or negative correlation for variance reduction, as discussed e.g. in Wilson (1985).

For c satisfying $F(c) < 1$, the residual moments, $E\{X^k | X \geq c\}$, when $\alpha = 0$, are

$$E\{X^k | X \geq c, \alpha = 0\} = \beta^k \left\{ (-\gamma)^k (\epsilon/\delta)^{k/2} \sum_{i=0}^k \frac{(-1)^i \binom{k}{i} (\gamma^{i\delta+1} - p_l^{i\delta+1})}{(i\delta+1) \gamma^{i\delta}} \right. \\ \left. + (1-\gamma)^k (\delta/\epsilon)^{k/2} \sum_{i=0}^k \frac{(-1)^i \binom{k}{i} (1-p_u)^{i\epsilon+1}}{(i\epsilon+1) (1-\gamma)^{i\epsilon}} \right\} / [1 - p_c]$$

for $k = 1, 2, \dots$, where $p_c \equiv F(c; \alpha = 0)$, $p_l \equiv \min\{\gamma, p_c\}$, and $p_u \equiv \max\{\gamma, p_c\}$.

The residual moments are derived in the Appendix.

The moments centered at α are obtained in the limit as $c \rightarrow -\infty$, in which case $p_l = 0$ and $p_u = \gamma$, yielding

$$\begin{aligned} E\{X^k | \alpha = 0\} &= \beta^k \{ -(-\gamma)^{k+1} (\epsilon/\delta)^{k/2} [\sum_{i=0}^k (-1)^i \binom{k}{i} / (1\delta + 1)] \\ &\quad + (1-\gamma)^{k+1} (\delta/\epsilon)^{k/2} [\sum_{i=0}^k (-1)^i \binom{k}{i} / (1\epsilon + 1)] \} \\ &= \beta^k \{ -(-\gamma)^{k+1} (\epsilon/\delta)^{k/2} \delta^{-1} B(k+1, \delta^{-1}) \\ &\quad + (1-\gamma)^{k+1} (\delta/\epsilon)^{k/2} \epsilon^{-1} B(k+1, \epsilon^{-1}) \} \end{aligned}$$

where $B(.,.)$ is the beta function, which can be evaluated numerically or, since one argument is integer, algebraically using

$$\begin{aligned} \delta^{-1} B(k+1, \delta^{-1}) &= \sum_{i=0}^k (-1)^i \binom{k}{i} / (1\delta + 1) \\ &= \delta^k k! / \prod_{i=0}^k (1\delta + 1) \end{aligned}$$

The specific four-parameter case of $\delta = \epsilon$ yields

$$E\{X^k | \alpha = 0, \delta = \epsilon\} = \beta^k [(1-\gamma)^{k+1} - (-\gamma)^{k+1}] \delta^{-1} B(k+1, \delta^{-1})$$

In the general five-parameter case, the mean is

$$E\{X\} = \alpha + \beta (\delta\epsilon)^{1/2} [((1-\gamma)^2/(\epsilon+1)) - (\gamma^2/(\delta+1))]$$

and when $\delta = \epsilon$

$$E\{X \mid \delta = \epsilon\} = \alpha + \beta [(1-2\gamma)\delta / (\delta+1)]$$

The mean is equal to the mode α when $\gamma = 1/2$ and $\delta = \epsilon$ or when $\gamma = [(\delta+1) - ((\delta+1)(\epsilon+1))^{1/2}] / (\delta-\epsilon)$ and $\delta \neq \epsilon$. Equivalently the mean and mode are equal if $\epsilon = [(1-\gamma)^2(\delta+1)/\gamma^2] - 1$ or $\delta = [\gamma^2(\epsilon+1)/(1-\gamma)^2] - 1$.

The variance can be calculated using $V\{X\} = E\{X^2 \mid \alpha = 0\} - E^2\{X \mid \alpha = 0\}$. In the specific four-parameter case of $\delta = \epsilon$, $V\{X \mid \delta = \epsilon\} = (\beta \delta)^2 [2\gamma(1-\gamma)(\delta-1) + 1] / [(2\delta+1)(\delta+1)^2]$.

3. PARAMETER DETERMINATION

To be useful in statistical modeling, a family of distributions requires methods for determining appropriate parameter values. Two contexts — (1) obtaining a set of specified properties such as moments and fractiles and (2) fitting to data — are discussed here. The mixed case of obtaining some properties and fitting to data is discussed briefly in Section 4.

3.1 Obtaining Specified Properties

Consider fitting a distribution to a set of specified properties, rather than to data. Situations in which no data are available often occur when modelling proposed systems that do not yet exist. The set of properties to be obtained arises from technological considerations and from expert (or nonexpert) opinion. The former is often lower and upper bounds and the latter is often low-order moments.

Four sets of properties are discussed here. In each case, appropriate parameter values can be found using closed-form formulas, sometimes combined with a single unidimensional search, to obtain any of several sets of properties. For each set of

properties, the corresponding parameter values are given below. Their derivation is straightforward and other sets can be derived as needed.

Here μ and σ^2 denote the mean and variance, a and b the lower and upper bounds, and m the mode. Also p_m is $F(m)$, (p_1, x_1) implies that $F(x_1) = p_1$ where $x_1 < \alpha$ and $p_1 < \gamma$, and (p_2, x_2) implies that $F(x_2) = p_2$ where $x_2 > \alpha$ and $p_2 > \gamma$.

I. Specify $a, m, p_m, (p_1, x_1)$, and (p_2, x_2) . Then

$$\alpha = m$$

$$\gamma = p_m$$

$$\delta = \ln((x_1 - a) / (\alpha - a)) / \ln(p_1 / \gamma)$$

Solve for ϵ recursively, by trial and error, or any unidimensional search

$$\text{procedure } \ln \epsilon = \delta [1 - ((1 - p_2) / (1 - \gamma))^\epsilon] ((1 - \gamma) / \gamma) (\alpha - a) / (x_2 - \alpha)$$

$$\beta = ((\alpha - a) / \gamma) (\delta / \epsilon)^{1/2}$$

II. Specify a, b, m, p_m , and (p_1, x_1) . Then

$$\alpha = m$$

$$\gamma = p_m$$

$$\delta = \ln((x_1 - a) / (\alpha - a)) / \ln(p_1 / \gamma)$$

$$\epsilon = \delta (\alpha - a) (1 - \gamma) / ((b - \alpha) \gamma)$$

$$\beta = ((\alpha - a) / \gamma) (\delta / \epsilon)^{1/2}$$

III. Specify a, b, μ , and m, p_m . Then

$$\alpha = m$$

$$\gamma = p_m$$

$$\beta = [(\alpha - a) (b - \alpha) / (\gamma (1 - \gamma))]^{1/2}$$

$$c = \beta \gamma / (\alpha - a)$$

Solve for ϵ in the quadratic equation

$$\epsilon^2 [c(\mu - \alpha) / \beta + \gamma^2 - c^2(1 - \gamma)^2] + \epsilon [(\mu - \alpha) (c^2 + 1) / (\beta c) + 2\gamma - 1]$$

$$+ (\mu - \alpha) / (\beta c) = 0$$

Then

$$\delta = \epsilon c^2$$

Here $c = (\delta/\epsilon)^{1/2}$ is an intermediate value.

IV. Specify a, μ, σ^2, m, p_m . Then

$$\alpha = m$$

$$\gamma = p_m$$

Solve for δ using

$$\begin{aligned} \sigma^2 + (\mu - \alpha)^2 = & \beta^2 \{ \gamma^3 (\epsilon/\delta) [(1/(2\delta+1)) - (2/(\delta+1)) + 1] \\ & + (1 - \gamma)^3 (\delta/\epsilon) [1 - (2/(\epsilon+1)) + (1/(2\epsilon+1))] \} \end{aligned}$$

where

$$\epsilon = \frac{(1-\gamma)^2}{\frac{\gamma(\mu-\alpha)}{\delta(\alpha-a)} + \frac{\gamma^2}{\delta+1}} - 1$$

and

$$\beta = ((\alpha-a)/\gamma) (\delta/\epsilon)^{1/2}$$

The four sets of properties are certainly not adequate for every situation. But given another set of properties, the user at most needs to solve five nonlinear equations for the five unknown values $\alpha, \beta, \gamma, \delta$ and ϵ . For most properties, however, algebraic simplification reduces the problem to one or two equations.

3.2 Parameter Estimation from Data

The more classic distribution fitting situation is to estimate parameter values from data x_1, x_2, \dots, x_n by assuming the data arise independently from some particular member of the family of distributions. Common methods include maximizing likelihood, matching moments, and minimizing a goodness-of-fit statistic such as the Kolmogorov-Smirnoff statistic, the chi-square, or a weighted squared deviation between the empirical and fitted distribution function. Ad hoc methods, as discussed in the next section, for fitting the five-parameter family work reasonably well, but a well-developed theory-based methodology has not yet been developed. As a first step, the five maximum likelihood equations, obtained by setting the derivatives of the likelihood function to zero, are stated here.

$$\alpha: (1-\delta) \sum_{i=1}^{n_1} [\gamma + z_i(\delta/\epsilon)^{1/2}]^{-1} = (1-\epsilon) \sum_{i=n_1+1}^n [(1-\gamma) - z_i(\epsilon/\delta)^{1/2}]^{-1} \quad (3)$$

$$\begin{aligned} \beta: n(\delta/\epsilon)^{1/2} &= (\delta-1) \sum_{i=1}^{n_1} [\gamma + z_i(\delta/\epsilon)^{1/2}]^{-1} z_i \\ &\quad - (\epsilon-1) \sum_{i=n_1+1}^n [(1-\gamma) - z_i(\epsilon/\delta)^{1/2}]^{-1} z_i \end{aligned} \quad (4)$$

$$\gamma: (\gamma-1)/\gamma = \frac{(1-\epsilon) \sum_{i=n_1+1}^n z_i / ((1-\gamma) - z_i(\epsilon/\delta)^{1/2})}{(1-\delta) \sum_{i=1}^{n_1} z_i / (\gamma + z_i(\delta/\epsilon)^{1/2})} \quad (5)$$

$$\begin{aligned}
\delta: \quad \delta = (\beta/n) \{ & -2 \sum_{i=1}^{n_1} \ln[1 + (z_i(\delta/\epsilon)^{1/2}/\gamma)] \\
& + (1-\delta) \sum_{i=1}^{n_1} [\gamma + z_i(\delta/\epsilon)^{1/2}]^{-1} z_i(\delta/\epsilon)^{1/2} \\
& + (1-\epsilon) \sum_{i=n_1+1}^n [(1-\gamma) - z_i(\epsilon/\delta)^{1/2}]^{-1} z_i(\delta/\epsilon)^{1/2} \} \quad (6)
\end{aligned}$$

$$\begin{aligned}
\epsilon: \quad \epsilon = (\beta/n) \{ & -2 \sum_{i=n_1+1}^n \ln[1 - z_i(\epsilon/\delta)^{1/2}/(1-\gamma)] \\
& + (\delta-1) \sum_{i=1}^{n_1} [\gamma + z_i(\delta/\epsilon)^{1/2}]^{-1} z_i(\epsilon/\delta)^{1/2} \\
& + (\epsilon-1) \sum_{i=n_1+1}^n [(1-\gamma) - z_i(\epsilon/\delta)^{1/2}]^{-1} z_i(\epsilon/\delta)^{1/2} \} \quad (7)
\end{aligned}$$

where $z_i = (x_i - \alpha)/\beta$ and n_1 is the number of x_i values less than α .

These equations are difficult to use for at least two reasons. The first is that n_1 is unknown and must be estimated indirectly via α . The second reason is inherent in all families of distributions for which the density function can be infinite at the end points of the distribution. With no additional constraints, the equations always lead to a U-shaped distribution ($\delta > 1, \epsilon > 1$) with end points at the minimum, $x_{(1)}$, and the maximum, $x_{(n)}$, since the likelihood function is then infinite. When the distribution is known not to have infinite density at the lower and/or upper bounds, the estimation procedure can be improved by specifying the lower bound, a , and/or the upper bound, b . A modified maximum likelihood algorithm is to solve the five nonlinear equations (5), (6), (7), $\beta = (b-a)/(\gamma(\epsilon/\delta)^{1/2} + (1-\gamma)(\delta/\epsilon)^{1/2})$, and $\alpha = a + \gamma \beta(\epsilon/\delta)^{1/2}$ using initial estimates for $\alpha, \beta, \gamma, \delta$, and ϵ obtained from one of the four sets of properties in

Section 3.1.

However, the authors' experience has been that solving the system of nonlinear equations often results in numerical complications. Minimizing the average squared deviation between the empirical and fitted distribution function, the approach taken by Öztürk and Dale (1985) for the generalized lambda distribution, is an approach worth investigating. The authors have obtained good fits by repeatedly using one of the four techniques of Section 3.1 for several input sets of values and choosing the parameters with the minimum associated goodness-of-fit value, as illustrated in the next section with grouped data and the χ^2 statistic.

4. NUMERICAL EXAMPLES

Two numerical examples illustrate the family of distributions in applied settings.

4.1 Modeling a Logic Network

In modeling a logic network, the propagation delay of each component is of interest. The Texas Instruments TTL Data Book (1973) gives the minimum time, typical time, and maximum time for a positive-NAND gate is 2, 7, and 15 nanoseconds, respectively. In a particular situation suppose the model calls for 1% of the gates to exceed the stated maximum of 15 ns and for the upper bound to be 20 ns. The four required properties (lower bound, mode, upper bound and (p_2, x_2)), which do not match any of the four cases in Section 3.1, are

$$\alpha = \text{mode} = 7 \text{ ns}$$

$$\alpha - \beta \gamma (\epsilon/\delta)^{1/2} = \text{lowerbound} = 2 \text{ ns}$$

$$\alpha + \beta (1-\gamma) (\delta/\epsilon)^{1/2} = \text{upperbound} = 20 \text{ ns}$$

and, substituting $(p_2, x_2) = (.99, 15\text{ns})$ into eq. (2),

$$15 = \alpha + \beta(1-\gamma)(\delta/\epsilon)^{1/2}[1 - ((1-.99)/(1-\gamma))^6]$$

Since these are only four properties and five parameters, one possibility is to set $\delta = \epsilon$.

The solution to these four equations is then closed form, yielding

$$\alpha = 7 \text{ ns}$$

$$\beta = b - a = 18 \text{ ns}$$

$$\gamma = 0.278$$

and

$$\delta = \epsilon = \ln[1 - (x_2 - \alpha)/(\beta(1-\gamma))] / \ln[(1-p_2)/(1-\gamma)] = 0.2233$$

4.2 Modeling Coefficient of Friction Data

Consider fitting to the 250 observed coefficients of friction in Hahn and Shapiro (1967, p. 219). The given frequency classes are

Table 1. The Hahn and Shapiro (1967) coefficient of friction data.

Coefficient of Friction	Percent of Observations
Less than 0.0150	0.4
0.015 - 0.020	3.6
0.020 - 0.025	12.0
0.025 - 0.030	17.6
0.030 - 0.035	23.2
0.035 - 0.040	18.0
0.040 - 0.045	11.6
0.045 - 0.050	6.8
0.050 - 0.055	3.6
0.055 - 0.060	1.6
0.060 or more	1.6

Set I of Section 3.1 was used based on forty-five sets of values for the all combinations of the properties

a: 0.0

(m, p_m): (.030, .336), (.0325, .452), (.035, .568)

(x₁, p₁): (.02, .04), (.025, .16), (.030, .336)

(x₂, p₂): (.035, .568), (.040, .748), (.045, .864), (.050, .932), (.055, .968)

The smallest χ^2 value, 4.387, was obtained when $a = 0.0$, $(m, p_m) = (.030, .336)$, $(x_1, p_1) = (.02, .04)$, and $(x_2, p_2) = (.045, .864)$. The corresponding parameter values are $\alpha = .03$, $\beta = .08115$, $\gamma = .336$, $\delta = .1905$, and $\epsilon = .2306$. The corresponding upper bound is $b = .07897$. As can be seen from the plot of the empirical distribution function and the fitted distribution function in Figure 3, the fit using this crude, nonoptimal method is good. The Johnson system S_U distribution fitted by matching moments yields a curve in Figure 3 that is difficult to distinguish visually from the five-parameter curve shown.

ACKNOWLEDGMENTS

This research was supported by Office of Naval Research contracts N00014-77-C-0425 and N00014-79-C-0832. We thank James R. Wilson for his helpful comments.

APPENDIX: Derivation of Residual Moments

For $k = 1, 2, \dots$, the k^{th} moment conditional on $\{X \geq c\}$ is derived here. For any $c \in (-\infty, \infty)$, by definition

$$E\{X^k \mid X \geq c, \alpha = 0\} = \begin{cases} \int_c^{\beta(1-\gamma)(\delta/\epsilon)^{1/2}} x^k f(x) dx / [1 - p_c] & \text{if } p_c < 1 \\ 0 & \text{if } p_c = 1 \end{cases}$$

Substituting for $f(x)$, we simplify the integral as follows.

$$\begin{aligned} & \int_c^{\beta(1-\gamma)(\delta/\epsilon)^{1/2}} x^k f(x) dx \\ &= \int_{p_l}^{\gamma} \{-\beta \gamma (\epsilon/\delta)^{1/2} [1 - (p/\gamma)^\delta]\}^k dp \\ & \quad + \int_{p_u}^1 \{\beta(1-\gamma)(\delta/\epsilon)^{1/2} [1 - ((1-p)/(1-\gamma))^\epsilon]\}^k dp \\ &= \beta^k \left\{ \gamma^k (\epsilon/\delta)^{k/2} \int_{p_l}^{\gamma} \left[\sum_{i=0}^k (-1)^i \binom{k}{i} (p/\gamma)^{(k-i)\delta} \right] dp \right. \\ & \quad \left. + (1-\gamma)^k (\delta/\epsilon)^{k/2} \int_{p_u}^1 \left[\sum_{i=0}^k (-1)^i \binom{k}{i} ((1-p)/(1-\gamma))^{\epsilon i} \right] dp \right\} \\ &= \beta^k \left\{ (-\gamma)^k (\epsilon/\delta)^{k/2} \sum_{i=0}^k \frac{(-1)^i \binom{k}{i} (\gamma^{\delta+1} - p_l^{\delta+1})}{(i\delta+1) \gamma^{i\delta}} \right. \\ & \quad \left. + (1-\gamma)^k (\delta/\epsilon)^{k/2} \sum_{i=0}^k \frac{(-1)^i \binom{k}{i} (1-p_u)^{\epsilon i+1}}{(i\epsilon+1) (1-\gamma)^{i\epsilon}} \right\} \end{aligned}$$

where $p_c \equiv F(c; \alpha = 0)$, $p_l \equiv \min\{\gamma, p_c\}$, and $p_u \equiv \max\{\gamma, p_c\}$, as defined in the text.

REFERENCES

- Burr, I.W. (1942), "Cumulative Frequency Functions," *Annals of Mathematical Statistics*, 13, 215-232.
- Burr, I.W. (1973), "Parameters for a General System of Distributions to Match a Grid of α_3 and α_4 ," *Communications in Statistics*, 2, 1-21.
- Dudewicz, E.J., Ramberg, J.S., and Tadikamalla, P.E. (1974), "A Distribution for Data Fitting and Simulation," *ASQC Transactions*, 28, 407-418.
- Hahn, G.J., and Shapiro, S.S. (1967), *Statistical Models in Engineering*, John Wiley and Sons, New York.
- Harter, H.L. (1967), "Maximum Likelihood Estimation of the Parameters of a Four-Parameter Gamma Population from Complete and Censored Samples," *Technometrics*, 9, 159-165.
- Hora, S.C. (1983), "Estimation of the Inverse Function for Random Variate Generation," *Communications of the ACM*, 26, 8, 590-594.
- Johnson, M.E., Tietjen, G.L., and Beckman, R.J. (1980), "A New Family of Probability Distributions with Applications to Monte Carlo Studies," *Journal of the American Statistical Association*, 75, 370, 276-279. (See correction, *JASA*, 76, 376, 1022.)
- Johnson, N.L., and Kotz, S. (1970), "Distributions in Statistics: Continuous Univariate Distributions-1," John Wiley and Sons, New York.
- Joiner, B.L., and Rosenblatt, J.R. (1971), "Some Properties of the Range in Samples from Tukey's Symmetric Lambda Distribution," *Journal of the American Statistical Association*, 66, 398-399.
- Kottas, J.F., and Lau H-S. (1979), "A Realistic Approach for Modeling Stochastic Lead Time Distributions," *IIE Transactions*, 11, 1, 54-60.
- Lau, H-S., and Zakl A. (1982), "The Sensitivity of Inventory Decisions to the Shape of Lead-Time-Demand Distribution," *IIE Transactions*, 14, 4, 265-271.

- Lurie, D., and Hartley, H.O. (1972), "Machine-Generation of Order Statistics for Monte Carlo Computations," *The American Statistician*, 26, 26-27.
- Öztürk, A., and Dale, R.E. (1985), "Least Squares Estimation of the Parameters of the Generalized Lambda Distribution," *Technometrics*, 27, 1, 81-84.
- Ramberg, J.S., and Schmelser, B.W. (1972), "An Approximate Method for Generating Symmetric Random Variables," *Communications of the ACM*, 15, 987-990.
- Ramberg, J.S., and Schmelser, B.W. (1974), "An Approximate Method for Generating Asymmetric Random Variables," *Communications of the ACM*, 17, 78-82.
- Ramberg, J.S., and Tadikamalla, P.R. (1978), "On the Generation of Subsets of Order Statistics," *Journal of Statistical Computation and Simulation*, 6, 239-241.
- Ramberg, J.S., Tadikamalla, P.R., Dudewicz, E.J., and Mykytka, E.F. (1979), "A Probability Distribution and Its Uses in Fitting Data," *Technometrics*, 21, 2, 201-214.
- Schmelser, B.W. (1977), "Methods for Modeling and Generating Probabilistic Components in Digital Computer Simulation When the Standard Distributions are Not Adequate: A Survey," *Proceedings of the Winter Simulation Conference*, Highland, Sargent, and Schmidt (eds.), 51-55.
- Schmelser, B.W. (1978a), "Generation of the Maximum (Minimum) Value in Digital Computer Simulation," *Journal of Statistical Computation and Simulation*, 8, 103-116.
- Schmelser, B.W. (1978b), "The Generation of Order Statistics in Digital Computer Simulation: A Survey," *Proceedings of the Winter Simulation Conference*, Highland, Nielsen, and Hull (eds.), IEEE, 137-140.
- Schmelser, B.W. and Deutsch, S.J. (1977), "A Versatile Family of Probability Distributions Suitable for Simulation," *AIIE Transactions*, 9, 176-182.
- Schucany, W.R. (1972), "Order Statistics in Simulation," *Journal of Statistical Computation and Simulation*, 1, 281-286.

- Stacy, E.W. (1962), "A Generalization of the Gamma Distribution," *Annals of Mathematical Statistics*, 33, 1187-1192.
- Stacy, E.W., and Mlham, G.A. (1965), "Parameter Estimation for a Generalized Gamma Distribution," *Technometrics*, 7, 349-358.
- Texas Instruments (1973), *TTL Databook*.
- Tukey, J.W. (1960), *Technical Report 36, Statistical Techniques Research Group*, Princeton University.
- Wilson, J.R. (1985), "Variance Reduction Techniques for Digital Simulation," *American Journal of Mathematical and Management Sciences*, in press.

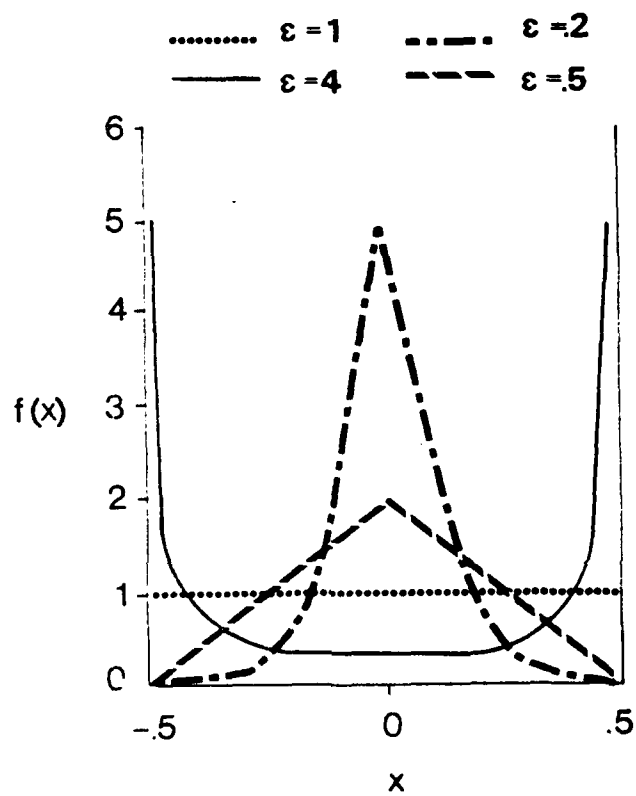


Figure 1. Symmetric densities for $\alpha = 0$, $\beta = 1$, $\gamma = 0.5$, and $\delta = \epsilon = 0.2, 0.5, 1, 4$.

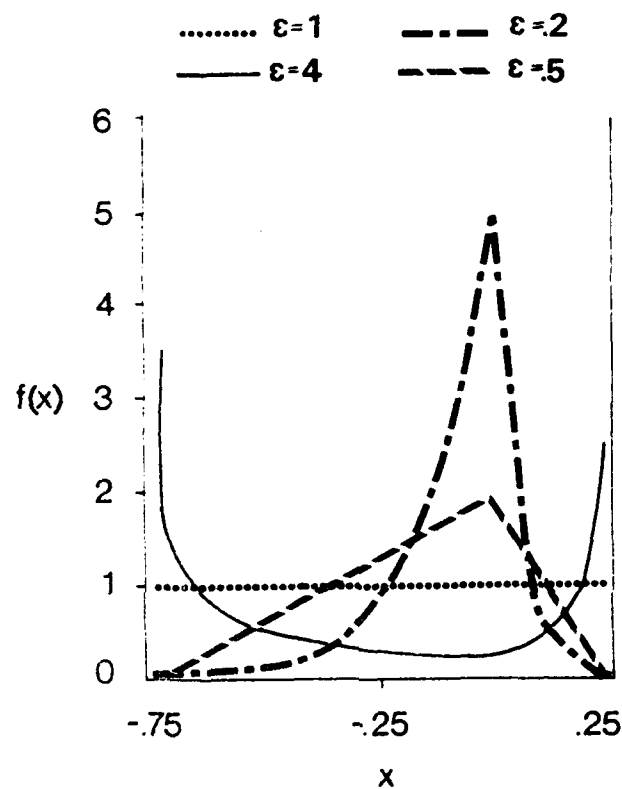


Figure 2. Asymmetric densities for $\alpha = 0$, $\beta = 1$, $\gamma = 0.75$, and $\delta = \epsilon = 0.2, 0.5, 1, 4$.

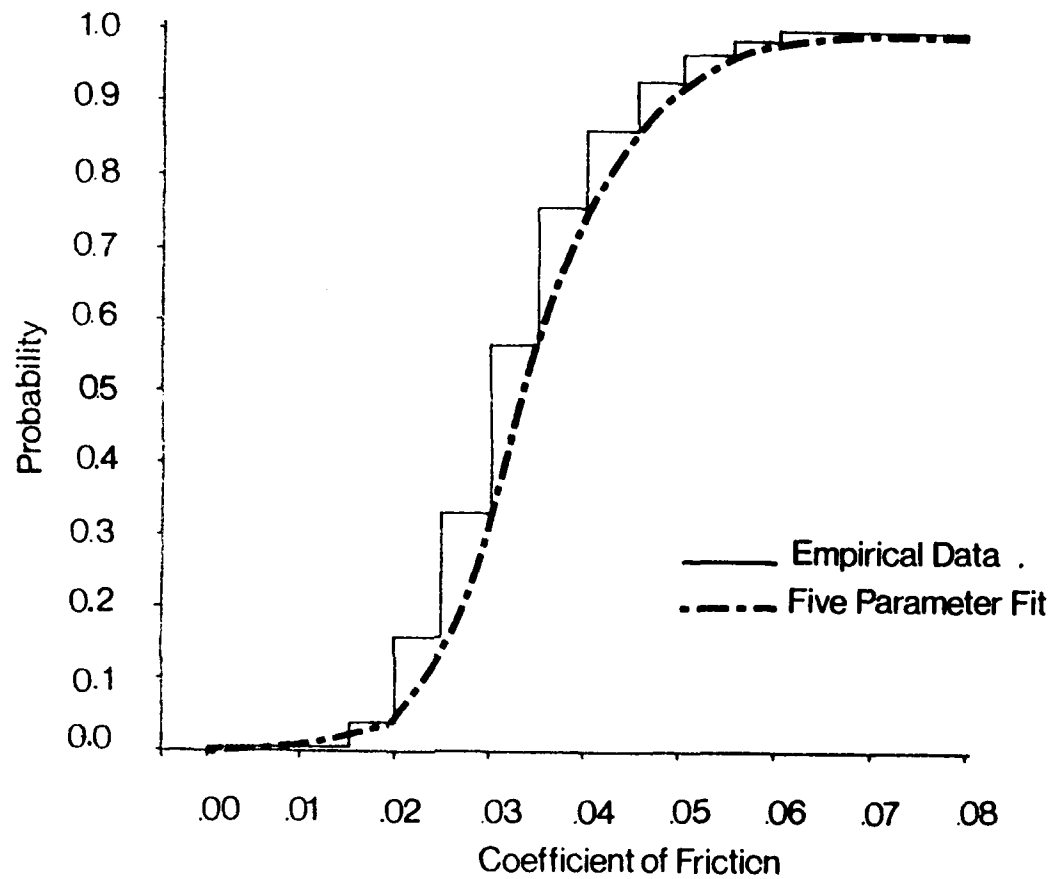


Figure 3. Five-parameter curve fitted to coefficient of friction data from Hahn and Shapfro (1987).

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 84-7	2. GOVT ACCESSION NO. AD A158645	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Five-Parameter Family of Probability Distributions		5. TYPE OF REPORT & PERIOD COVERED Research Memorandum
		6. PERFORMING ORG. REPORT NUMBER 84-7, revision 1
7. AUTHOR(s) Bruce W. Schmeiser Ram Lal		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0832
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Industrial Engineering Purdue University West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS M042-477
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research -- Code 411 800 North Quincy Street Arlington, VA 22217		12. REPORT DATE August 1985
		13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release: Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Input modeling Statistical modeling Monte Carlo Random variate generation Process generation Simulation		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A tractable, five-parameter family of continuous, unimodal probability distributions is developed. Special cases include the Bernoulli trial, uniform, power series, exponential, triangular, and Laplace (double exponential) distributions. The family has closed-form density function, cumulative distribution function, inverse distribution function, hazard function, and residual moments. Statistical properties, parameter determination, and random variate generation are discussed. Two numerical examples are given.		

DD FORM 1473

EDITION OF 1 NOV 63 IS OBSOLETE
S/N 0102-014-6601Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

10-85

DTIC